

On the Exponential Probability Bounds for the Bernoulli Random Variables

VLADIMIR NIKULIN*

Department of Mathematics, University of Queensland
Brisbane, Australia

Abstract

We consider upper exponential bounds for the probability of the event that an absolute deviation of sample mean from mathematical expectation p is bigger comparing with some ordered level ε . These bounds include 2 coefficients $\{\alpha, \beta\}$. In order to optimize the bound we are interested to minimize linear coefficient α and to maximize exponential coefficient β . Generally, the value of linear coefficient α may not be smaller than one. The following 2 settings were proved: 1) $\{1, 2\}$ in the case of classical discreet problem as it was formulated by Bernoulli in the 17th century, and 2) $\{1, \frac{2}{1+\varepsilon^2}\}$ in the general discreet case with arbitrary rational p and ε . The second setting represents a new structure of the exponential bound which may be extended to continuous case.

1 Introduction

Let X, X_1, X_2, \dots be a sequence of independent and identically distributed random variables $\mathbf{Pr}(X = 1) = p, \mathbf{Pr}(X = 0) = 1 - p$.

Ya. Bernoulli proved [14] that

$$\mathbf{Pr}(|\frac{1}{n} \sum_{i=1}^n X_i - p| > \varepsilon) \leq \frac{1}{1 + C},$$

where

$$C = \min(\frac{1}{s-1} \exp(\xi_1 \log \frac{r+1}{r}), \frac{1}{r-1} \exp(\xi_2 \log \frac{s+1}{s})),$$

*Email: vnikulin.uq@gmail.com

$$\xi_1 = \frac{k(r+1)+s}{r+s+1}, \xi_2 = \frac{k(s+1)+r}{r+s+1},$$

$p = \frac{r}{r+s}, n = k(r+s), \varepsilon = \frac{1}{r+s}$, and k, r, s are arbitrary natural numbers.

As far as $\xi_1, \xi_2 \geq \frac{k}{r+s} = n\varepsilon^2$ we can make conclusion that

$$C = \min\left(\frac{1}{s-1} \exp(n\varepsilon^2 \log \frac{r+1}{r}), \frac{1}{r-1} \exp(n\varepsilon^2 \log \frac{s+1}{s})\right).$$

Therefore, we can formulate the law of large numbers

$$\lim_{n \rightarrow \infty} \mathbf{Pr}\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - p\right| > \varepsilon\right) = 0,$$

where r and s are assumed to be fixed and $k \rightarrow \infty$.

Markov [10], [12] considered case of arbitrary n, p and ε . Uspensky [16] extended results of Markov further and derived the first exponential bound

$$\mathbf{Pr}\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - p\right| \geq \varepsilon\right) \leq \alpha \exp\{-\beta \varepsilon^2 n\} \quad (1)$$

with coefficients $\alpha = 2, \beta = 0.5$. Additional and more detailed historical notes may be found in the Section “Existing Inequalities” [1].

Hoeffding [8] developed methods of [16], [4] and proved generally that $\alpha = \beta = 2$. Note that similar exponential bounds for the empirical distribution functions were presented in [5] and [13].

In the Section 3 we prove that the following values of the constants $\alpha = 1$ and $\beta = 2$ may be used in the bound (1) in the discrete case formulated and considered by Ya. Bernoulli. It is demonstrated in the Section 3.1 that value of α can not be smaller than one. The following Section 3.2 proves one-sided inequalities using methods and results of the Theorem 1.

Section 4 introduces a new structure of the exponential bound in the general discrete case with arbitrary rational p and ε . The best bound in asymptotical sense corresponds to the bigger value of the exponential coefficient. However, as it is discussed in the Section 4.1, the value of linear coefficient is also very important because of the practical reasons.

Section 5 represents an extension of the methods and results of the Section 4 to the continuous case.

Section 5.1 considers particular application of the bounds to the normalized sum of random variables. It is demonstrated that all 3 types of bounds are asymptotically equivalent. In this way, advantage of the proposed bounds comparing with Hoeffding’s bound is absolute.

Note that using results of [2] and [3] we can extend bounds for Bernoulli random variables to the case of arbitrary bounded random variables.

Also, we note paper [9] where similar exponential bounds were constructed for Markov chains. Generally, exponential bounds proved to be very effective in order to define required size of the sample in order to ensure proper quality of estimation, see, for example, [11, 7, 6, 15, 17].

2 Main lemma and definitions

We will use essentially different approach comparing with [8] and [4]. This technique is based on the properties of convex (concave) functions applied to the binomial coefficients. The following Lemma formulates the core of the methods which are employed in the Sections 3, 4 and 5.

Lemma 1 *Suppose that $\varphi(t), 1 \leq t \leq n$, is a function with continuous second derivative where n is any natural number. Then,*

$$\sum_{j=1}^n \varphi(j) \geq n\varphi\left(\frac{n+1}{2}\right) \quad (2)$$

if $\varphi'(t) \leq 0, \varphi''(t) \geq 0, 1 \leq t \leq n$ (convex case);

$$\sum_{j=1}^n \varphi(j) \geq n \frac{\varphi(1) + \varphi(n)}{2} \quad (3)$$

if $\varphi'(t) \geq 0, \varphi''(t) \leq 0, 1 \leq t \leq n$ (concave case).

Proof: The following representations are valid

$$\sum_{j=1}^n \varphi(j) = \begin{cases} \sum_{j=0}^{m-1} [\varphi(1+j) + \varphi(n-j)] & \text{if } n = 2m; \\ \sum_{j=0}^{m-1} [\varphi(1+j) + \varphi(n-j)] + \varphi(m+1) & \text{if } n = 2m+1, \end{cases}$$

where m is a natural number. We obtain required bounds combining above equation with

$$\varphi(1+j) + \varphi(n-j) \geq 2\varphi\left(\frac{n+1}{2}\right) \geq 2\varphi(m+1), 1 \leq j \leq m-1,$$

in the convex case, and with

$$2\varphi(m+1) \geq \varphi(1+j) + \varphi(n-j) \geq \varphi(1) + \varphi(n), 1 \leq j \leq m-1,$$

in the concave case. ■

The following notations will be used below

$$\begin{aligned}\mathcal{P}^0 &:= \mathbf{Pr}\left(\frac{1}{n} \sum_{i=1}^n X_i = p\right) = \binom{n}{m} p^m (1-p)^{n-m}, \\ \mathcal{P}_1^+ &:= \mathbf{Pr}\left(0 < \frac{1}{n} \sum_{i=1}^n X_i - p \leq \varepsilon\right), \quad \mathcal{P}_2^+ := \mathbf{Pr}\left(\varepsilon < \frac{1}{n} \sum_{i=1}^n X_i - p\right), \\ \mathcal{P}_1^- &:= \mathbf{Pr}\left(0 < p - \frac{1}{n} \sum_{i=1}^n X_i \leq \varepsilon\right), \quad \mathcal{P}_2^- := \mathbf{Pr}\left(\varepsilon < p - \frac{1}{n} \sum_{i=1}^n X_i\right),\end{aligned}$$

where $m = kr$.

Assuming that $n = k(r + s)$ we form groups of binomial probabilities of the equal size k

$$\begin{aligned}S_j &:= \sum_{\ell=m-kj}^{m-1-(j-1)k} \binom{n}{\ell} p^\ell (1-p)^{n-\ell}, j = 1, \dots, r \text{ (left groups)}, \\ Z_j &:= \sum_{\ell=m+1+(j-1)k}^{m+kj} \binom{n}{\ell} p^\ell (1-p)^{n-\ell}, j = 1, \dots, s \text{ (right groups)}.\end{aligned}$$

Then, we consider relations of the corresponding binomial coefficients from the neighbor groups

$$A(j) := \frac{\binom{n}{m-j}}{\binom{n}{m-k-j}} = \prod_{v=1}^k \frac{n-m+j+v}{m-k-j+v}, j = 1, \dots, k(r-1), \quad (4a)$$

$$B(j) := \frac{\binom{n}{m+j}}{\binom{n}{m+k+j}} = \prod_{v=1}^k \frac{m+j+v}{n-m-k-j+v}, j = 1, \dots, k(s-1), \quad (4b)$$

where $s, r \geq 2$.

By definition

$$\mathbf{Pr}\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - p\right| > \varepsilon\right) = \mathcal{P}_2^+ + \mathcal{P}_2^- = \sum_{j=2}^s Z_j + \sum_{j=2}^r S_j. \quad (5)$$

Remark 1 Note that $\mathcal{P}_2^+ = 0$ if $s = 1$, and $\mathcal{P}_2^- = 0$ if $r = 1$.

3 Bernoulli problem

We exclude from consideration the trivial cases: $p = 0$ and $p = 1$, and assume that $0 < p < 1$. As it will be demonstrated below the task of estimation of \mathcal{P}_2^+ is easier comparing with estimation of \mathcal{P}_2^- if $p > 0.5$. On the other hand, the task of estimation of \mathcal{P}_2^- is easier comparing with estimation of \mathcal{P}_2^+ if $p < 0.5$. As far as the problem is symmetrical, we assume that $p \geq 0.5$.

Theorem 1 *Suppose that $p = \frac{r}{r+s}$, $n = k(r+s)$, $\varepsilon = \frac{1}{r+s}$, and k, r, s are arbitrary natural numbers. Then,*

$$b \cdot Z_{j+1} \leq Z_j, j = 1, \dots, s-1,$$

where $b := \exp \{2\varepsilon^2 n\}$. Therefore,

$$(b-1)\mathcal{P}_2^+ \leq \mathcal{P}_1^+. \quad (6)$$

Proof: By (4b), $B(j)$ is an increasing function of $j \geq 1$. Therefore,

$$\lambda_j := \frac{Z_j}{Z_{j+1}} \geq B((j-1)k+1) \left(\frac{1-p}{p} \right)^k = \prod_{v=1}^k \frac{m+(j-1)k+1+v}{n-m-jk-1+v} \left(\frac{s}{r} \right)^k$$

or

$$\log \lambda_j \geq k \log \frac{s}{r} + \sum_{v=1}^k \varphi(j, v), j = 1, \dots, s-1,$$

where

$$\varphi(j, v) = \log \frac{m+(j-1)k+1+v}{n-m-jk-1+v}$$

is a convex function of v because $r \geq s$ or $m \geq n-m$, and according to (2)

$$\begin{aligned} \lambda_j &\geq \left(\frac{s(m+jk-0.5k+1.5)}{r(n-m-jk+0.5k-0.5)} \right)^k \\ &\geq \left(\frac{s(m+jk-0.5k+2)}{r(n-m-jk+0.5k)} \right)^k = (1+\delta)^k, \end{aligned} \quad (7)$$

where

$$\delta = \frac{n(j-0.5) + 2s}{kr(s-j+0.5)}.$$

Using general inequality

$$\log(1+\delta) \geq \frac{2\delta}{2+\delta} \quad \forall \delta \geq 0, \quad (8)$$

we obtain

$$\begin{aligned}\log \lambda_j &\geq \frac{2k(n(j-0.5)+2s)}{2(s-j+0.5)kr+n(j-0.5)+2s} \\ &\geq \frac{2kn(j-0.5)}{2srk+(j-0.5)(n-2kr)}, j=1, \dots, s-1.\end{aligned}\quad (9)$$

Furthermore, based on the following properties

$$2srk+(j-0.5)(n-2rk)=k[r(2s-j+0.5)+s(j-0.5)]>0, \quad 2kr \geq n,$$

we have

$$\log \lambda_j \geq \frac{n(j-0.5)}{sr}.$$

Then, using relations $r = \frac{p}{\varepsilon}, s = \frac{1-p}{\varepsilon}$ we transform above inequality to the required form

$$\lambda_j \geq \exp \left\{ \frac{n\varepsilon^2(2j-1)}{2p(1-p)} \right\} \geq \exp \{2n\varepsilon^2(2j-1)\}, j=1, \dots, s-1. \quad (10)$$

Above equation (10) completes proof of the Theorem. ■

Theorem 2 Suppose that $p = \frac{r}{r+s}, n = k(r+s), \varepsilon = \frac{1}{r+s}$, and k, r, s are arbitrary natural numbers. Then,

$$b \cdot S_{j+1} \leq S_j, j=1, \dots, r-1.$$

Therefore,

$$(b-1)\mathcal{P}_2^- \leq \mathcal{P}_1^-. \quad (11)$$

Proof: By (4a), the coefficient $A(j)$ is an increasing function of $j \geq 1$. Therefore,

$$q_j := \frac{S_j}{S_{j+1}} \geq A((j-1)k+1) \left(\frac{p}{1-p} \right)^k \quad (12)$$

or

$$\log q_j \geq k \log \frac{r}{s} + \sum_{v=1}^k \psi(j, v),$$

where $\psi(j, v) = \log \frac{n-m+(j-1)k+1+v}{m-jk-1+v}$.

The following condition

$$k(r-s-1) \geq 2$$

give us a guarantee that the function $\psi(1, v)$ is concave.

Suppose that $r = s + 2$ and $k \geq 2$ or $r \geq s + 3$ and $k \geq 1$. Then, by (3) we have

$$\begin{aligned} \log q_j &\geq \log q_1 \geq k \log \frac{r}{s} + \frac{k}{2} \log \frac{(n-m+2)(n-m+k+1)}{(m-k)(m-1)} \\ &\geq \frac{k}{2} \log \frac{r(s+1)}{s(r-1)} = \frac{k}{2} \log \left(1 + \frac{r+s}{s(r-1)} \right) \geq \frac{k(r+s)}{2s(r-1)+r+s} \quad (13) \\ &= \frac{n\varepsilon^2}{2p(1-p) + \varepsilon(2p-1)} \geq 2n\varepsilon^2 \end{aligned}$$

if $4p(1-p) + 2\varepsilon(2p-1) \leq 1$ or $\varepsilon \leq \frac{1-4p(1-p)}{2(2p-1)} = p - 0.5$. Thus, $\varepsilon = \frac{1}{r+s} \leq \frac{r}{r+s} - 0.5$ or $r \geq s + 2$.

The case $r = s + 2, k = 1$ is easy to consider:

$$\log q_1 = \log \frac{(s+2)^2}{s(s+1)} = \log \left(1 + \frac{3s+4}{s(s+1)} \right) \geq \frac{6s+8}{2s^2+5s+4} \geq \frac{\gamma}{2(s+1)} = \gamma n\varepsilon^2$$

or

$$(12 - 2\gamma)s^2 + (28 - 5\gamma)s + 16 - 4\gamma \geq 0.$$

Above inequality is valid $\forall s \geq 1$ if $\gamma \leq 4$.

Now, we consider remaining case $r = s + 1$ (the case $r = s$ was considered already in the first part of the proof because the problem is symmetrical). Then, $\psi(j, v)$ is convex as a function of v , and, by (2),

$$q_j \geq \left(\frac{(s+1)(ks+jk+1.5-0.5k)}{s(ks-jk-0.5+1.5k)} \right)^k = (1+\delta)^k, \quad (14)$$

where

$$\delta = \frac{2jsk+2s-ks+jk+1.5-0.5k}{s(ks-jk-0.5+1.5k)}.$$

Using inequality (8), we obtain

$$\log q_j \geq \frac{2k(2jsk+2s-ks+jk+1.5-0.5k)}{2s^2k+s+2sk+jk+1.5-0.5k} \geq \frac{\gamma k}{2s+1} = \gamma n\varepsilon^2 \quad (15)$$

or

$$\begin{aligned} (8jk+8-4k-2k\gamma)s^2 + (8jk-4k+10-(2k+1)\gamma)s + 2jk+3-k \\ -(jk+1.5-0.5k)\gamma \geq 0. \end{aligned}$$

Above inequality is valid for $\gamma \leq 2$ assuming that $j, s \geq 1$. ■

Corollary 1 *The following upper bound is valid*

$$\Pr(|\frac{1}{n} \sum_{i=1}^n X_i - p| > \varepsilon) < \exp \{-2\varepsilon^2 n\}, \quad (16)$$

where $p = \frac{r}{r+s}$, $n = k(r+s)$, $\varepsilon = \frac{1}{r+s}$, and k, r, s are arbitrary natural numbers.

Proof: It follows from the Theorems 1 and 2 that

$$\sum_{j=2}^r S_j \leq \frac{S_1}{b-1}, \quad \sum_{j=2}^s Z_j \leq \frac{Z_1}{b-1}. \quad (17)$$

The following representation is valid

$$\Pr(|\frac{1}{n} \sum_{i=1}^n X_i - p| > \varepsilon) = \frac{\sum_{j=2}^r S_j + \sum_{j=2}^s Z_j}{\mathcal{P}^0 + \sum_{j=1}^r S_j + \sum_{j=1}^s Z_j}.$$

It easy to see that $f(x) = \frac{x}{C+x}$ is an increasing function of x , where C is a positive constant. Therefore,

$$\Pr(|\frac{1}{n} \sum_{i=1}^n X_i - p| > \varepsilon) < \frac{\frac{S_1}{b-1} + \frac{Z_1}{b-1}}{S_1 + Z_1 + \frac{S_1}{b-1} + \frac{Z_1}{b-1}} = \frac{1}{b} = \exp(-2n\varepsilon^2).$$

■

3.1 Optimal value of the linear coefficient

Clearly, value of the linear coefficient $\alpha = 2$ in the exponential bound may not be regarded as an optimal. The following Proposition 1 demonstrates that $\alpha \geq 1$.

Proposition 1 *Suppose that $0 < p < 1$. Then, we can construct sequence of deviations $\varepsilon(n) \xrightarrow{n \rightarrow \infty} 0$ so that $\varepsilon^2 n \xrightarrow{n \rightarrow \infty} 0$ and $\mathcal{P}_2^- + \mathcal{P}_2^+ \xrightarrow{n \rightarrow \infty} 1$.*

Proof: By definition $\mathcal{P}^0 = \binom{n}{m} p^m (1-p)^{n-m}$. The following relation is valid

$$\log \binom{n}{m} = \sum_{j=1}^{n-m} \log \frac{m+j}{j}.$$

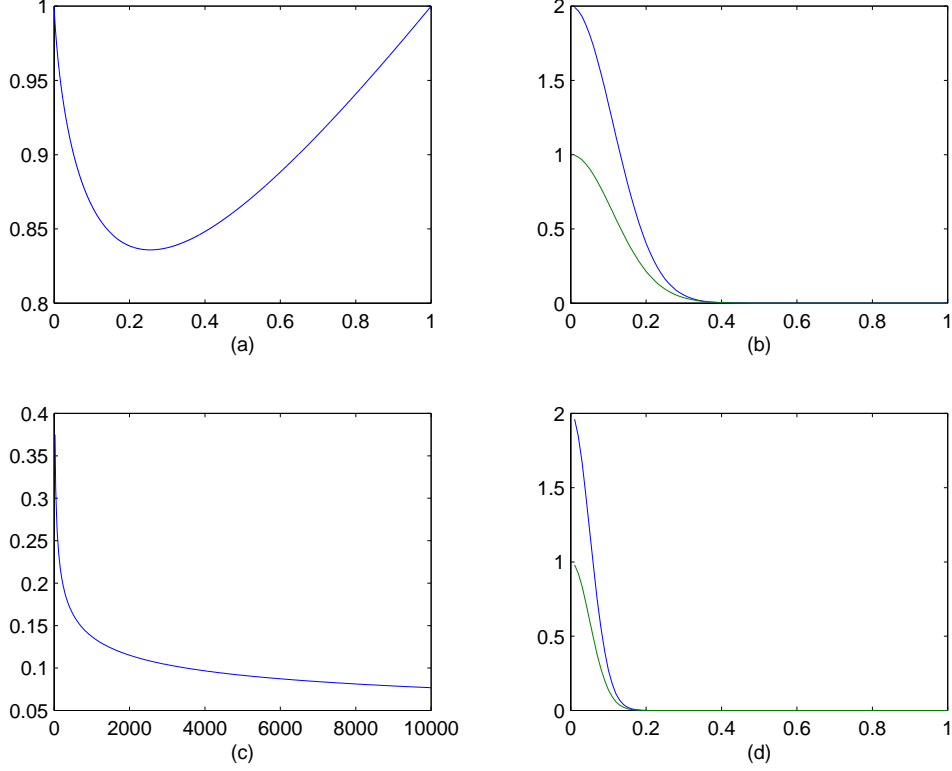


Figure 1: a) $f(p) = p^p(1+p)^{1-p}, 0 \leq p \leq 1$; c) function $\mu(n)$ which is defined in (33); bounds (31) and (32) as a function of ε where b) $n = 20$, d) $n = 100$.

By Lemma 1,

$$\log \binom{n}{m} \leq (n-m) \log \frac{1+n+m}{1+n-m} \leq (n-m) \log \frac{n+m}{n-m} = (n-m) \log \frac{1+p}{1-p}.$$

Respectively,

$$\mathcal{P}^0 \leq (p^p(1+p)^{1-p})^n,$$

where $0 < p^p(1+p)^{1-p} < 1 \quad \forall p : 0 < p < 1$. Figure 1(a) illustrates graph of the function $f(p) = p^p(1+p)^{1-p}, 0 \leq p \leq 1$.

Suppose that $\varepsilon = \frac{1}{2n}$. Then, \mathcal{P}_1^+ and \mathcal{P}_1^- will be empty sets, and $\mathcal{P}_2^- + \mathcal{P}_2^+ \xrightarrow[n \rightarrow \infty]{} 1$ as far as $\mathcal{P}^0 \xrightarrow[n \rightarrow \infty]{} 0$. At the same time, by definition, $\varepsilon^2 n \xrightarrow[n \rightarrow \infty]{} 0$. ■

Corollary 2 *Coefficient α in the bound (1) may not be smaller than 1 for the arbitrary $n \geq 2$ and $\varepsilon > 0$.*

3.2 One-sided inequalities

In this Section we will use again the property that left groups of binomial probabilities may be estimated more effectively if $p < 0.5$. Analogously, right groups of binomial probabilities may be estimated more effectively if $p > 0.5$.

Proposition 2 *Suppose that $p = \frac{r}{r+s}$, $n = k(r+s)$, $\delta = \nu\varepsilon = \frac{\nu}{r+s}$, where k, r, s and ν are arbitrary natural numbers. Then,*

$$\Pr\left(\frac{1}{n} \sum_{i=1}^n X_i - p > \delta\right) \leq \frac{1}{2} \exp\left(-\frac{n\delta^2}{2p(1-p)}\right)$$

if $r \geq s$, ($p \geq 0.5$);

$$\Pr\left(p - \frac{1}{n} \sum_{i=1}^n X_i > \delta\right) \leq \frac{1}{2} \exp\left(-\frac{n\delta^2}{2p(1-p)}\right)$$

if $r \leq s$, ($p \leq 0.5$).

Proof: Combining the equality

$$\sum_{i=j}^{\nu} (2i-1) = \nu^2 - (j-1)^2, j = 1, \dots, \nu,$$

and (10) we derive that

$$Z_j \geq Z_{\nu+1} \exp\left\{\frac{n\varepsilon^2(\nu^2 - (j-1)^2)}{2p(1-p)}\right\}, j = 1, \dots, \nu,$$

and

$$\sum_{j=1}^{\nu} Z_j \geq Z_{\nu+1} \sum_{j=1}^{\nu} \exp\left\{\frac{n\varepsilon^2(\nu^2 - (j-1)^2)}{2p(1-p)}\right\}. \quad (18)$$

Next, we use the same bound (10) in the opposite direction

$$Z_j \geq Z_{j+1} \exp\left\{\frac{n\varepsilon^2(2\nu+1)}{2p(1-p)}\right\}, j = \nu+1, \dots, s-1,$$

and

$$\sum_{j=\nu+1}^s Z_j \leq Z_{\nu+1} \left(1 - \exp\left\{-\frac{n\varepsilon^2(2\nu+1)}{2p(1-p)}\right\}\right)^{-1}. \quad (19)$$

Again, we make an assumption $r \geq s$ or $p \geq 0.5$ without loss of generality, and construct upper bound for the conditional probability

$$\begin{aligned} \Pr\left(\frac{1}{n} \sum_{i=1}^n X_i - p > \delta \mid \frac{1}{n} \sum_{i=1}^n X_i > p\right) &= \frac{\sum_{j=\nu+1}^s Z_j}{\sum_{j=1}^s Z_j} \\ &\leq \left(\left(1 - \exp\left(-\frac{n\varepsilon^2(2\nu+1)}{2p(1-p)}\right)\right) \sum_{j=1}^{\nu} \exp\left(\frac{n\varepsilon^2(\nu^2 - (j-1)^2)}{2p(1-p)}\right) + 1 \right)^{-1}, \end{aligned} \quad (20)$$

where the last inequality may be regarded as a consequence of (18) and (19).

Then, we simplify (20) using properties $\nu^2 - (j-1)^2 - 2\nu - 1 \leq \nu^2 - j^2$, $j = 1, \dots, \nu$,

$$\Pr\left(\frac{1}{n} \sum_{i=1}^n X_i - p > \delta \mid \frac{1}{n} \sum_{i=1}^n X_i > p\right) \leq \exp\left\{-\frac{n\delta^2}{2p(1-p)}\right\}. \quad (21)$$

As a next step we prove that

$$\Pr\left(\frac{1}{n} \sum_{i=1}^n X_i \leq p \mid p \geq 0.5\right) \geq \Pr\left(\frac{1}{n} \sum_{i=1}^n X_i > p \mid p \geq 0.5\right).$$

Let us consider relations of the binomial probabilities which are symmetrical against central point $m = kr$

$$\begin{aligned} \frac{\binom{n}{m-i+1}}{\binom{n}{m+i}} \left(\frac{1-p}{p}\right)^{2i-1} &= \prod_{j=1}^{2i-1} \frac{(m-i+1+j)(n-m)}{(n-m-i+j)m} \\ &\geq \left(\frac{m+1}{m}\right)^{2i-1} > 1, i = 1, \dots, n-m, \end{aligned}$$

where we used (2). Therefore,

$$\Pr\left(\frac{1}{n} \sum_{i=1}^n X_i \geq p\right) = \Pr\left(\frac{1}{n} \sum_{i=1}^n X_i \geq p \mid p \geq 0.5\right) \leq 0.5. \quad (22)$$

By the way, using identical method, one can prove

$$\Pr\left(\frac{1}{n} \sum_{i=1}^n X_i \leq p \mid p \leq 0.5\right) \leq 0.5.$$

The final upper bound follows from (21) and (22) applied to the Bayesian formulae

$$\begin{aligned} & \Pr\left(\frac{1}{n} \sum_{i=1}^n X_i - p > \delta\right) \\ &= \Pr\left(\frac{1}{n} \sum_{i=1}^n X_i \geq p\right) \cdot \Pr\left(\frac{1}{n} \sum_{i=1}^n X_i - p > \delta \mid \frac{1}{n} \sum_{i=1}^n X_i > p\right). \end{aligned}$$

■

4 General discreet case

In this section we assume that p and ε may be represented by positive rational numbers with denominator as a number of observations in the sample. Respectively, we will cover all possible empirical values of the sample mean as an estimator of the probability p . Note that the role of the parameter k will be different here comparing with previous section.

Theorem 3 *Suppose that $p = \frac{m}{n} \geq 0.5$, $\varepsilon = \frac{k}{n}$, and k, m, n are arbitrary natural numbers with condition $k, m < n$. Then,*

$$b \cdot Z_{j+1} \leq Z_j, j = 1, \dots, s-1,$$

where $b := \exp\{2\varepsilon^2 n\}$. Therefore,

$$(b-1)\mathcal{P}_2^+ \leq \mathcal{P}_1^+. \quad (23)$$

Proof: The following upper bound may be obtained similarly to (7)

$$\log \lambda_j \geq k \log \frac{(n-m)(m+0.5k+2)}{m(n-m-0.5k)}, j = 1, \dots, s-1, \quad (24)$$

where $2m \geq n$ and $p + \frac{\varepsilon}{2} < 1$ (else $\mathcal{P}_2^+ = 0$).

It follows from (24) that

$$\log \lambda_j \geq k \log(1 + \delta) \quad (25)$$

where $\delta = \frac{0.5kn+2n-2m}{m(n-m-0.5k)}$.

By (8), we obtain

$$\frac{k(kn+4n-4m)}{2nm-2m^2-km+0.5kn+2n-2m} \geq \frac{\gamma k^2}{n} = \gamma \varepsilon^2 n$$

or

$$k + 4 - 4p \geq \gamma k \left[2p(1 - p) + \frac{\varepsilon}{2}(1 - 2p) \right] + 2\gamma\varepsilon(1 - p).$$

Above inequality is valid $\forall k \geq 1$ and $\forall p \geq 0.5$ if $\gamma \leq 2$. ■

The following Theorem 4 introduces a new structure of the exponential bound as a main result of this paper.

Theorem 4 *Suppose that $p = \frac{m}{n} \geq 0.5$, $\varepsilon = \frac{k}{n}$, and k, m, n are arbitrary natural numbers with conditions $2 \leq k < n, m < n$. Then,*

$$\exp \left\{ \frac{2\varepsilon^2 n}{1 + \varepsilon^2} \right\} \cdot S_{j+1} \leq S_j, j = 1, \dots, r - 1.$$

Therefore,

$$(\exp \left\{ \frac{2\varepsilon^2 n}{1 + \varepsilon^2} \right\} - 1) \mathcal{P}_2^- \leq \mathcal{P}_1^- \quad (26)$$

(the case $k = 1$ is covered by the Theorem 2 with stronger result).

Proof: Similar to (13) and assuming that $2m > n + k + 2$ we derive

$$\log q_j \geq k \log \frac{m}{n - m} + \frac{k}{2} \log \frac{(n - m + 2)(n - m + k + 1)}{(m - k)(m - 1)}, \quad (27)$$

where $j = 1, \dots, r - 1, m > k \geq 1$ (else $\mathcal{P}_2^- = 0$).

The following relations are valid

$$\frac{(n - m + 2)(n - m + k + 1)}{(m - k)(m - 1)} \left(\frac{m}{n - m} \right)^2 \geq \frac{m(n - m + k + 1)}{(m - k)(n - m)} = 1 + \delta,$$

where $\delta = \frac{m + kn}{(m - k)(n - m)}$.

By (8), we obtain

$$\log q_j \geq \frac{k(m + kn)}{2m(n - m + k) + m - kn} \geq \frac{\gamma k^2}{n} = \gamma n \varepsilon^2$$

or

$$n(m + kn) \geq \gamma k [2m(n - m + k) + m - kn]$$

or

$$p + k \geq \gamma k \left[2p(1 - p + \varepsilon) + \frac{p}{n} - \varepsilon \right] = \gamma k [2p(1 - p) + \varepsilon(2p - 1)] + \gamma \varepsilon p.$$

Above inequality take place if

$$\gamma \leq \frac{2}{1 + \varepsilon^2}. \quad (28)$$

Let us consider remaining case $n \leq 2m \leq n + k + 2$. Similar to (7) we have

$$q_j \geq \left(\frac{m(n - m + 1.5 + 0.5k)}{(n - m)(m - 0.5k - 0.5)} \right)^k = (1 + \delta)^k, \quad (29)$$

where

$$\delta = \frac{m + 0.5kn + 0.5n}{nm - m^2 + 0.5(km + m - kn - n)}.$$

By (8), we obtain

$$\frac{2k(m + 0.5kn + 0.5n)}{2nm - 2m^2 + km + 2m - 0.5n(k + 1)} \geq \frac{\gamma k^2}{n} = \gamma \varepsilon^2 n$$

or

$$k + 2p + 1 \geq \gamma k(2p(1 - p) + 0.5\varepsilon(2p - 1)) + \gamma \varepsilon(2p - 0.5).$$

Above inequality take place if

$$\gamma \leq \frac{8}{4 + \varepsilon^2}. \quad (30)$$

Note that condition (30) is less restrictive comparing with (28). ■

Corollary 3 *The following upper bound is valid*

$$\Pr\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - p\right| > \varepsilon\right) < \exp\left\{-\frac{2\varepsilon^2 n}{1 + \varepsilon^2}\right\}, \quad (31)$$

where $p = \frac{m}{n}$, $\varepsilon = \frac{k}{n}$, and k, m, n are arbitrary natural numbers with conditions $2 \leq k < n, n \leq 2m < 2n$.

Proof is similar to the Corollary 1 : we obtain required value of the parameter β as $\min\{2, \frac{2}{1+\varepsilon^2}, \frac{8}{4+\varepsilon^2}\}$ where first value follows from Theorem 3, second and third values follow from (28) and (30).

4.1 Relations between bounds

The bound (31) improves Hoeffding's bound

$$\Pr\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - p\right| > \varepsilon\right) \leq 2 \exp\{-2\varepsilon^2 n\} \quad (32)$$

if

$$0 < \varepsilon^2 \leq \mu^2(n) = \frac{\phi}{2} + \sqrt{\phi \left(1 + \frac{\phi}{4}\right)}, \quad \phi := \frac{\log 2}{2n}. \quad (33)$$

Table 1: The following parameters were used in this example: $n = 33, m = 15, k = 1, \dots, 14, \varepsilon = \frac{k}{n}$, second column represents real probability of absolute deviation of the sample mean from $p = \frac{m}{n}$, next two columns represent corresponding bound (31) and Hoeffding's bound (32).

ε	True probability	Bound (31)	Hoeffding's Bound (32)
0.0606	0.600713	0.78542	1.569446
0.0909	0.382439	0.582175	1.159157
0.1212	0.220522	0.38456	0.758396
0.1515	0.114271	0.227376	0.43955
0.1818	0.052796	0.120996	0.225672
0.2121	0.021571	0.058319	0.102638
0.2424	0.007724	0.025643	0.041352
0.2727	0.0024	0.010366	0.014758
0.3030	0.00064	0.003884	0.004666
0.3333	0.000145	0.00136	0.001307
0.3636	0.000027	0.000449	0.000324
0.3939	0.000004	0.000141	0.000071
0.4242	0.000001	0.000042	0.000014
0.4545	0	0.000012	0.000002

Therefore,

$$\lim_{n \rightarrow \infty} n^{\frac{1}{4}} \mu(n) = \left(\frac{\log 2}{2} \right)^{\frac{1}{4}}.$$

Figure 1(c) illustrates behavior of the function μ .

Remark 2 *It follows from (33) that the number of observations must be big enough in order to ensure advantage of the Hoeffding's bound for the fixed deviation parameter ε :*

$$n > \frac{\log 2}{2\varepsilon^4}.$$

As a direct consequence, the bound (31) will be so small:

$$\exp \left\{ -\frac{2\varepsilon^2 n}{1 + \varepsilon^2} \right\} < \exp \left\{ -\frac{\log 2}{\varepsilon^2(1 + \varepsilon^2)} \right\}$$

that any further improvement may not be regarded as a significant. Figures 1(b)(d) illustrate above fact with relatively small numbers of observations $n = 20$ and $n = 100$.

5 Continuous Case

Suppose that p and ε are arbitrary numbers:

$$0 < p < 1; \quad (34)$$

$$\frac{1}{n} < \varepsilon \leq \min\{p, 1 - p\}. \quad (35)$$

Remark 3 *The special case $\varepsilon \leq \frac{1}{n}$ may be considered easily. In the case $\varepsilon > p$ or $\varepsilon > 1 - p$ we will have simplified cases because we will need to approximate only one probability of deviation \mathcal{P}_2^+ or \mathcal{P}_2^- .*

We define a central point np which is not necessarily integer. We denote by 1) $h(n)$ - number of integer numbers in the left group $[np - n\varepsilon, np]$; 2) $g(n)$ - number of integer numbers in the right group $[np, np + n\varepsilon]$. All remaining left groups will have h integers with only one possible exception as a last group. Symmetrically, all remaining right groups will have g integers with only one possible exception as a last group.

Let us denote by m the smallest integer in $[np, np + n\varepsilon]$.

According to the construction

$$0 \leq \tilde{p} - p \leq \frac{1}{n}; \quad (36)$$

$$|\tilde{\varepsilon}_i - \varepsilon| \leq \frac{1}{n}; \quad (37)$$

$$\theta := \max\left\{\frac{\varepsilon}{\tilde{\varepsilon}_1}, \frac{\varepsilon}{\tilde{\varepsilon}_2}\right\} \leq \frac{n\varepsilon}{n\varepsilon - 1}, \quad (38)$$

where $\tilde{p} := \frac{m}{n}$, $\tilde{\varepsilon}_1 := \frac{h(n)}{n}$, $\tilde{\varepsilon}_2 := \frac{g(n)}{n}$.

Theorem 5 *Suppose that conditions (34) and (35) are valid. Then,*

$$\Pr\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - p\right| > \varepsilon\right) \leq \exp\left\{\varepsilon \cdot \varphi(n, \varepsilon) - \frac{2\varepsilon^2 n}{1 + \varepsilon^2}\right\}, \quad (39)$$

where

$$\varphi(n, \varepsilon) := \frac{\frac{n\varepsilon(1+\varepsilon^2)}{n\varepsilon-1} + \varepsilon(2 + 6\varepsilon + \frac{9}{2n})}{(1 + \varepsilon^2)(1 + \varepsilon^2 + \frac{1}{n}(1 + 3\varepsilon + \frac{9}{4n}))}. \quad (40)$$

Proof: Again, we make an assumption $p \geq 0.5$. Similar to (24) we obtain

$$\log \lambda_1 \geq g \log \frac{(1-p)(m+0.5g+1.5)}{p(n-m-0.5g-0.5)} = g \log (1+\delta), \quad (41)$$

where $\delta = \frac{m+1.5+0.5g-p-pn}{pn-mp-0.5p-0.5gp}$.

By (8), we obtain

$$\frac{2(m+1.5+0.5g-p-pn)}{pn-2mp-gp-2p+m+1.5+0.5g} \geq \frac{\gamma g}{n}$$

or

$$\tilde{\varepsilon}_2 + 2(\tilde{p} - p) + \frac{3-2p}{n} \geq \gamma \tilde{\varepsilon}_2 \left(p - 2\tilde{p}p - \tilde{\varepsilon}_2 p - \frac{2p}{n} + \tilde{p} + 0.5\tilde{\varepsilon}_2 + \frac{3}{2n} \right).$$

Assuming that $\gamma \leq 2$ we simplify above inequality according to (36) by making it stronger

$$\tilde{\varepsilon}_2 + \frac{3-2p}{n} \geq \gamma \tilde{\varepsilon}_2 (2p(1-p) + \frac{1}{n}(1.5-2p)).$$

As a next step we can split above inequality into 2 inequalities

$$\tilde{\varepsilon}_2 \geq 2\gamma \tilde{\varepsilon}_2 p(1-p); \quad \frac{3-2p}{n} \geq \frac{\gamma \tilde{\varepsilon}_2}{n}(1.5-2p),$$

which are valid if

$$\gamma \leq 2. \quad (42)$$

Now, we consider approximation in the left direction according to the previous framework (27)

$$\begin{aligned} \log q_1 &\geq \frac{h}{2} \log \left(\left(\frac{p}{1-p} \right)^2 \frac{(n-m+2)(n-m+1+h)}{(m-h)(m-1)} \right) \\ &\geq \frac{h}{2} \log \left(\frac{p(n-m+1+h)}{(1-p)(m-h)} \right) = \frac{h}{2} \log (1+\delta), \end{aligned}$$

where $\delta = \frac{pn+p-m+h}{m-mp-h+hp}$.

By (8), we obtain

$$\frac{pn+p-m+h}{m-2mp-h+2hp+pn+p} \geq \frac{\gamma h}{n} = \gamma \tilde{\varepsilon}_1$$

or

$$\tilde{\varepsilon}_1 + p - \tilde{p} + \frac{p}{n} \geq \gamma \tilde{\varepsilon}_1 \left[\tilde{p}(1-2p) + p - \tilde{\varepsilon}_1(1-2p) + \frac{p}{n} \right].$$

Table 2: Values of the function $\exp \{\varepsilon \varphi(n, \varepsilon)\}$.

$n \setminus \varepsilon$	0.02	0.05	0.1	0.2	0.3	0.35
100	1.0412	1.0697	1.1436	1.3737	1.7612	2.0357
1,000	1.0221	1.0582	1.1336	1.3652	1.7566	2.0349
100,000	1.0211	1.0572	1.1326	1.3643	1.7561	2.0348

Again, we simplify above inequality according to (36) and (37)

$$\tilde{\varepsilon}_1 - \frac{1}{2n} \geq \gamma \tilde{\varepsilon}_1 \left[2p(1-p) + \varepsilon(2p-1) + \frac{3p-1}{n} \right]. \quad (43)$$

Therefore,

$$\gamma \leq \frac{2 - \frac{\theta}{n\varepsilon}}{1 + \varepsilon^2 + \frac{1}{n}(1 + 3\varepsilon + \frac{9}{4n})}. \quad (44)$$

The last remaining case corresponds directly to (29).

$$\log q_1 \geq h \log \left(\frac{p(n-m+1.5+0.5h)}{(1-p)(m-0.5h-0.5)} \right) = h \log(1+\delta)$$

where $\delta = \frac{pn+p-m+0.5(h+1)}{m-pm+0.5(hp+p-h-1)}$.

By (8), we obtain

$$\tilde{\varepsilon}_1 + 2(p - \tilde{p}) + \frac{2p+1}{n} \geq \gamma \tilde{\varepsilon}_1 \left[\tilde{p}(1-2p) + p + \frac{\tilde{\varepsilon}_1}{2}(2p-1) + \frac{2p-0.5}{n} \right].$$

We simplify above inequality according to (36) and (37)

$$\tilde{\varepsilon}_1 \geq \gamma \tilde{\varepsilon}_1 \left[2p(1-p) + \frac{\varepsilon(2p-1)}{2} + \frac{3p}{n} - \frac{1}{n} \right] \quad (45)$$

which is valid for any

$$0 < \gamma \leq 2 \left(\frac{4 + \varepsilon^2}{4} + \frac{1}{n} \left(1 + \frac{3\varepsilon}{2} + \frac{9}{4n} \right) \right)^{-1}. \quad (46)$$

Finally, we derive required asymptotical relation as a consequence of the conditions (42), (44) and (46) where condition (44) is the most restrictive.

■

5.1 Asymptotical Bounds for Normalized Sum of Bernoulli Random Variables

Let us denote by F_n distribution function of the normalized sum of Bernoulli random variables

$$\eta_n = \frac{\sum_{i=1}^n X_i - p}{\sqrt{np(1-p)}}. \quad (47)$$

According to the Central Limit Theorem,

$$\lim_{n \rightarrow \infty} F_n(t) = \Phi(t),$$

where $\Phi(t)$ is a standard normal distribution function. Respectively, it appears to be reasonable to use random variable (47) as a test case in order to compare different bounds.

Proposition 3 *The bounds (16), (31) and (39), as an upper bounds for the following probability $\Pr(|\eta_n| > t)$, are equivalent asymptotically and equal to*

$$\alpha \exp \{-2t^2 p(1-p)\}, \quad (48)$$

where $\alpha = 1$, and $0 \leq t, p \leq 1$.

Proof: We have

$$\Pr(|\eta_n| > t) = \Pr\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - p\right| > t_n\right),$$

where $t_n = t \sqrt{\frac{p(1-p)}{n}}$.

Then, we insert t_n into (16), (31) or (39), and take \lim if $n \rightarrow \infty$. As a result, we obtain required formulae (48). ■

Remark 4 *Using Hoeffding's bound we will obtain the same asymptotical structure (48), but with $\alpha = 2$.*

6 Concluding remarks

The Proposition 1 proves that value of the linear coefficient $\alpha = 1$ can not be improved. Taking this fact as a base point we established a new structure of the exponential bound (31). Figures 1(a),(c), (d) and Table 1 demonstrate advantages of the bound (31) against Hoeffding's bound if value of

ε is small enough. The above Section 5.1 demonstrates additional arguments in support of the proposed bounds, and these arguments cover not only Bernoulli random variables. The area of applications of the proposed methods may be extended further, for example, we can consider arbitrary bounded random variables or uniform metric for the empirical distributions.

The paper represents a fresh look at the ideas and methods which Ja. Bernoulli proposed in the 17th century. As it was demonstrated in the Section 3 the probability of deviation in the classical Bernoulli case may be bounded using standard structure of the exponential bound with optimal linear coefficient $\alpha = 1$. It was the first step of this research which was completed in 1987 shortly after the 1st World Congress of the Bernoulli Society where the author purchased book [14].

References

- [1] G. Bennett. "Probability Inequalities for the Sum of Independent Random Variables." *Journal of the American Statistical Association*, vol. 57, pp. 33-45, 1962.
- [2] V. Bentkus. "A remark on Bernstein, Prokhorov, Bennett, Hoeffding, and Talagrand inequalities." *Lithuanian Mathematical Journal*, vol. 42(3), pp. 262-269, 2002.
- [3] V. Bentkus. "On Hoeffding's inequalities." *The Annals of Probability*, vol. 32(2), pp. 1650-1673, 2004.
- [4] H. Chernoff. "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations." *Annals of Mathematical Statistics*, vol. 23, pp. 493-507, 1952.
- [5] A. Dvoretzky, J. Kiefer and J. Wolfowitz. "Asymptotic Minimax Character of the Sample Distribution Function and of the Classical Multinomial Estimator." *Annals of Mathematical Statistics*, vol. 27, pp. 642-669, 1956.
- [6] J. Gama, P. Medas and R. Rocha. "Forest trees for on-line data." *ACM Symposium on Applied Computing*, pp. 632-636, 2004.
- [7] Z. Heszberger, J. Zatoryi, J. Biro and T. Henk, "Efficient bounds for bufferless statistical multiplexing." *IEEE Globecom, San Francisco, USA*, pp. 641-645, 2000.

- [8] W. Hoeffding. "Probability Inequalities for Sums of Bounded Random Variables." *Journal of the American Statistical Association*, vol. 58, pp. 13-30, 1963.
- [9] C. Leon and F. Perron. "Optimal Hoeffding bounds for discrete reversible Markov chains." *The Annals of Applied Probability*, vol. 14(2), pp. 958-970, 2004.
- [10] Yu. V. Linnik. "Selected papers of A.A. Markov: On the problem of Yakob Bernoulli (1914)." *Academy of Science, USSR*, 1951.
- [11] G. Lueker. "Exponentially small bounds on the expected optimum of the partition and subset sum problems." *Random Structures and Algorithms*, 12, pp. 51-62, 1998.
- [12] A. A. Markov. *Calculus of Probabilities*, Gosizdat, 1924.
- [13] P. Massart. "The Tight Constant in the Dvoretzky-Kieffer-Wolfowitz Inequality." *Annals of Probability*, vol. 18(3), pp. 1269-1283, 1990.
- [14] Yu. V. Prochorov. "Ya. Bernoulli: On the Law of Large Numbers (1713)." *Nauka*, 1986.
- [15] P. Ravikumar and J. Lafferty. "Variational Chernoff bounds for graphical models." *20th Conference on Uncertainty in Artificial Intelligence*, 2004.
- [16] J. V. Uspensky. "Introduction to Mathematical Probability." *New York; London: Mc Graw-Hill*, 1937.
- [17] F. Yeh and M. Gallagher. "An empirical study of Hoeffding racing for model selection in k-nearest neighbor classification." *IDEAL 2005, LNCS 3578, Springer-Verlag, M. Gallagher and J. Hogan and F. Maire*, pp. 220-227, 2005.